# Structural Analysis of Wikigraph to Investigate Quality Grades of Wikipedia Articles

Anamika Chhabra
Department of CSE,
Indian Institute of Technology Ropar
India
anamika.chhabra@iitrpr.ac.in

Shubham Srivastava
Department of CSE,
Punjab Engineering College, Chandigarh
India
shubhamsrivastava1909@gmail.com

S. R. S. Iyengar
Department of CSE,
Indian Institute of Technology Ropar
India
sudarshan@iitrpr.ac.in

Poonam Saini
Department of CSE,
Punjab Engineering College, Chandigarh
India
poonamsaini@pec.edu.in

## ABSTRACT

The quality of Wikipedia articles is manually evaluated which is time inefficient as well as susceptible to human bias. An automated assessment of these articles may help in minimizing the overall time and manual errors. In this paper, we present a novel approach based on the structural analysis of Wikigraph to automate the estimation of the quality of Wikipedia articles. We examine the network built using the complete set of English Wikipedia articles and identify the variation of network signatures of the articles with respect to their quality. Our study shows that these signatures are useful for estimating the quality grades of un-assessed articles with an accuracy surpassing the existing approaches in this direction. The results of the study may help in reducing the need for human involvement for quality assessment tasks.

## CCS CONCEPTS

• **Information systems** → **Wikis**; • **Human-centered computing** → *Empirical studies in collaborative and social computing*.

## KEYWORDS

Wikipedia, Quality estimation, Wikigraph, Network analysis

**ACM Reference Format:**
Anamika Chhabra, Shubham Srivastava, S. R. S. Iyengar, and Poonam Saini. 2021. Structural Analysis of Wikigraph to Investigate Quality Grades of Wikipedia Articles. In *Companion Proceedings of the Web Conference 2021 (WWW '21 Companion), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3442442.3452345

## 1 INTRODUCTION

Due to the freely-accessible nature of Wikipedia, there is a significant disparity in the number and expertise of contributors as well as how they coordinate on different articles [9, 11, 18]. This results in a considerable variation in the quality level of Wikipedia articles. Owing to the popularity of Wikipedia, readers, as well as editors, have always been interested in knowing about the quality details of its articles [1, 2, 22]. This has instigated research into coming up with different ways of computing the quality of Wikipedia articles [12, 21, 27, 28, 30]. Currently, Wikipedia declares the quality of its articles in the form of a quality grade assigned to each article[1]. This grading system is based on a letter scheme, particularly reflecting the factual completeness of an article. Excluding the grades pertaining to the *List-class* articles[2], the grades roughly in the order of descending quality are: *Featured Articles (FA)*, *A-class*, *Good Articles (GA)*, *B-class*, *C-class*, *Start* and *Stub*. FA articles represent the class of comprehensive articles that follow professional standards of writing and presentation. A-class represents articles that provide a complete description of the topic. GA articles seem to have no obvious problems and approach the quality of a professional encyclopedia. B and C-class articles, although decent enough, require some more work in terms of completeness. Start-class articles lack completeness as well as sufficient references. The lowest quality grade is stub class containing articles that provide just a little more content than a dictionary definition.

The assignment of grades helps a reader identify good-quality articles and also assists an editor trace poor-quality articles that require attention [16]. However, the procedure to assign these grades is non-trivial. It employs a voting-based evaluation procedure where any user can nominate an article for the respective quality grade. The nominated articles are further scrutinized either by editors working for *Wiki-projects* or by independent editors. Apart from being a laborious and time-consuming process, it is inherently subjective and thus prone to human bias. This is because editors have their own opinion on the perceived quality of an article based on their expertise and domain knowledge [11]. Moreover, Wikipedia articles are dynamic in nature where the existing content triggers more content into the articles [10]. Therefore, the assessment must be carried out periodically in order to keep the ratings relevant and up-to-date with the latest content. This requires editors to


*WWW '21 Companion, April 19–23, 2021, Ljubljana, Slovenia*

ACM ISBN 978-1-4503-8313-4/21/04.
https://doi.org/10.1145/3442442.3452345

[1]https://en.wikipedia.org/wiki/Wikipedia:Content_assessment
[2]*List-class* articles are non-prose articles and contain lists of other articles.

spend a substantial amount of their time and effort on the articles' assessment rather than utilizing it to curate knowledge and refinement.

Researchers have attempted to estimate the quality grades of Wikipedia articles through automated [4, 27, 30] or semi-automated ways [21]. Most of the existing works on estimating the articles' quality have focused on features pertaining to the articles' content or their contributors' properties. However, our work shows that the properties of these articles' underlying network provide a valuable set of features to train the machine learning models. We investigate the Wikigraph network where nodes are Wikipedia articles which are linked through *internal links*[3] that connect Wikipedia articles with each other. Further, as compared to prior studies involving a subset of articles, we examine the network created by the complete set of Wikipedia articles, thus providing a comprehensive investigation. We compute various node-specific properties of the articles on the basis of their position in Wikigraph and observe how their values vary with respect to their quality grades. We find that the node-specific network properties of the articles correlate with their quality grades. We use this observation to train our classification model, which helps automate the process of estimating the quality of the articles with respect to the entire spectrum of the quality grades ranging from stub to FA. The ability to gauge the quality through automated means may help the editors expend their time and effort in working on the articles' content, thereby reducing their requirement in administrative activities.

## 2 RELATED WORK

The research in the direction of examining quality grades of Wikipedia articles may be divided into two streams: one where the binary classification of the articles into low and high-quality articles is performed, and the other where the prediction across the entire spectrum of quality grades, i.e., six (or seven) classes is carried out.

In the direction of binary classification, Stvilia et al. [27] use different article characteristics - computed based on parameters such as the number of users, reverts, broken links, internal links, readability, etc - to predict whether an article belongs to the featured set or random set. A later study showed that a simple measure such as the articles' word count can roughly help in predicting article quality [4]. Yet another study used life-cycle based metrics to predict high and low-quality articles [30]. These metrics differentiate between transient and persistent contributions, where transient contributions are those that are quickly reverted soon enough while the persistent contributions tend to stay in the article. Using the network analysis of Wikipedia, Brandes et al. [5] examine the network of collaboration between Wikipedia editors-referred to as *edit network*- and compare the controversial articles with non-controversial featured articles. The authors find that the structural network parameters are correlated with the quality labels of the articles. Another work examines article-editor networks of Wikipedia and develops models to rank the articles [21]. The authors propose a combination of manual evaluation and automatic evaluation as a useful solution for the articles' quality assessment. Apart from these, a few other measures adopted for examining article quality in Wikipedia are bonding and diversity among the editors [23], examining the evolution of articles over time [33], team characteristics [3] and NLP features of the articles [12].

A few prior works have also used the multi-class classification to predict one of the six (or seven) quality grades for Wikipedia articles. Warncke et al. [28, 29] use features such as the number of headings, links, references, images, etc, and classify the articles into seven quality grades. Another well-known model for quality prediction ORES (Object Revision Evaluation Service) [15] uses these features as well as includes more actionable features and divides into six quality grades while improving accuracy. Another work by Dang et al. [12] uses Doc2vec and deep neural networks to predict across six quality grades. A very recent work by Raman et al. [13] uses features obtained from the revision history networks of articles to obtain classification across six classes. As compared to all these works, our method obtains a better accuracy. A comparison of these is reported in Section 4.

Using Wikigraph - the kind of network that we are using in our work - only a few studies have been performed, although not towards investigating the quality grades. Buriol et al. [7] studied temporal details of Wikigraph. They examined the growth properties of the number of articles, visitors, and editors (exponential), size of articles (linear), and the number of links per article (slow linear). Zlatic et al. [34] studied Wikigraphs of different language Wikipedias and observed that they exhibit similar growth patterns with respect to their degree distributions, topology, reciprocity, clustering, assortativity, path lengths, and triad significance profiles, etc.

## 3 PROPOSED APPROACH

A broad outline of our proposed approach is provided in Figure 1. The numbers on the arrows represent the sequencing of different operations. The first step involves gathering the relevant data pertaining to Wikigraph. From this data, node-specific features of all the articles are computed. These features, along with the quality grades of the articles that have already been assessed by the Wikipedia community form the processed data set (Step-2). This data set is then used to train the prediction model (Step-3). For an input article whose quality grade is to be predicted, its node-specific features form the input data (Step-4) that is fed into the trained model (Step-5) and the quality grade is predicted (Step-6).

### 3.1 Building the data set

For obtaining the network of all the articles of English Wikipedia, we downloaded the data from the publicly available data dump of Wikipedia[4]. The data obtained from here contained all Wikipedia pages, including talk pages, user pages, etc. Different Wikipedia pages - such as article pages, talk pages, and user pages, etc - in this data are differentiated by 34 different namespaces. Our study focuses on main/article pages that belong to the namespace 0. The article pages that are 'redirects' to other articles are not considered. The resulting network consists of $6,007,492$ articles (including $5,980,643$ assessed and $493,561$ un-assessed articles) and $453,663,058$ links among them. This study focuses on predicting the quality grades of the articles belonging to seven quality grades, i.e., Start, Stub, C, B, Good articles (GA), A, and Featured Articles

[3] https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Linking

[4] https://dumps.wikimedia.org/enwiki/latest/

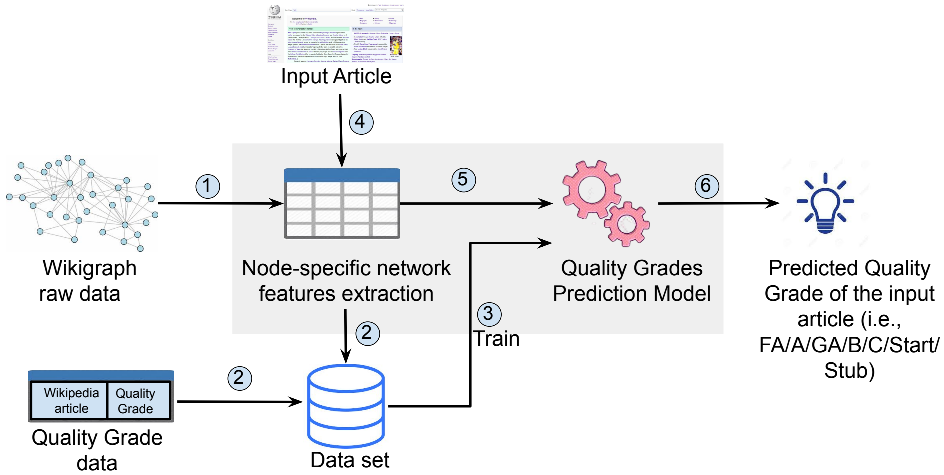



**Figure 1: Schematic diagram of the proposed approach. The numbers on the arrows represent the sequencing of different operations.**

(FA). The articles from classes 'List' ($n = 256,087$) and 'Featured List (FL)' ($n = 2,020$) are not examined due to their non-traditional layout and a different creation method. The number of articles belonging to each of these classes is shown in Table 1.

**Table 1: Number of articles belonging to each quality grade as accessed in March, 2019 dump**

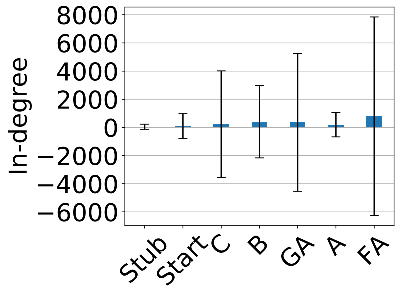
| Stub | Start | C | B | GA | A | FA |
|---|---|---|---|---|---|---|
| 3,390,074 | 1,826,843 | 332,133 | 131,091 | 33,420 | 2,109 | 6,866 |

## 3.2 Computing Node-specific properties of Wikigraph

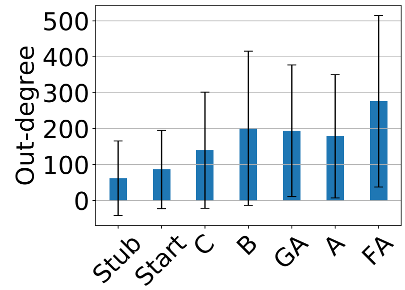
An article exhibits certain properties by virtue of its position in the Wikigraph. This results from the kind of connections it makes with the other articles, which in turn are affected by its quality. Therefore, an examination of the different node-specific properties of the articles may be able to provide us comparative details of their quality. Following are the properties that we examined for all the articles of the Wikigraph:

(1) **In-degree and out-degree**: Top quality articles are comprehensive and broad in their coverage as compared to the articles belonging to low-quality classes. Therefore, in the Wikigraph, top quality articles are expected to have more in-degree as well as out-degree.
(2) **Betweenness Centrality:** Betweenness centrality [14] captures the extent to which the nodes stand on the shortest paths between each other. Due to more connectivity, good quality articles are expected to fall on a high number of shortest paths between the nodes in the network. Hence we compute their betweenness centrality.
(3) **Katz-centrality:** Katz Centrality [17] considers all the walks between nodes than only computing the shortest paths. It may thus help in providing an alternate measure of influence of the articles in the Wikigraph.
(4) **PageRank:** PageRank algorithm [25] takes into account the number as well as quality of links to a node and estimates how important the node is. It uses the assumption that important nodes are more likely to receive a higher volume of links from other important nodes. In WikiGraph, more comprehensive articles are likely to be more connected to the articles that are themselves highly-connected. Therefore, we compute the PageRank of the articles.
(5) **Clustering-coefficient**: Clustering Coefficient measures the extent to which the neighbors of a given node are connected to each other. Given the variation in the number of neighbors as well as the connections among them, this property may provide patterns with respect to the quality grades.
(6) **Hub and Authority Scores:** We use HITS (Hyperlink-Induced Topic Selection) algorithm [19] which computes two values for each node: *Hub* score and *Authority* Score. The nodes with a high authority score are authoritative sources of information on a topic while those with a high hub score compile authoritative sources. In the context of WikiGraph, these two scores may provide interesting details with respect to the quality grades of the articles.
(7) **Hirsch-index (H-index):** The H-index [6] of a node in a directed network is defined to be the maximum value $h$ such that there exist at least $h$ neighbors having in-degree at least $h$. Therefore, this parameter improves upon simple measure of degree and hence may provide additional details of the articles that a given article is connected to.
(8) **Shell Number:** We use *k-shell decomposition* method [8] that recursively prunes nodes from the network as per their degree, starting from the nodes with lowest degree. In each pass $i$, the nodes $u$ having degree $d(u) \leq i$ are pruned until there are no nodes of degree $i$ left in the network. The nodes pruned in pass $i$ may be visualized as being kept in bucket $i$. The nodes in the buckets form the shells of the network, where lower shells form the periphery while the

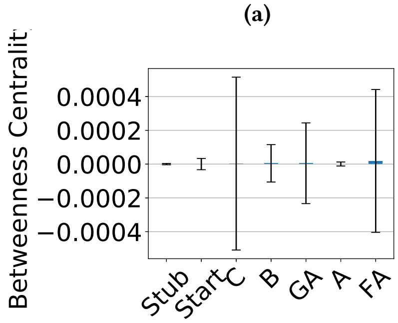
**Figure 2: Average and SD of (a) In-degree (b) Out-degree (c) Betweenness Centrality (d) Katz Centrality (e) Page Rank (f) Clustering Coefficient (g) Hub Score (h) Authority Score values, (i) H-index and (j) Shell numbers of the articles belonging to each quality grade. The values mostly exhibited increasing or decreasing patterns for the properties, thus indicating their usefulness for prediction.**



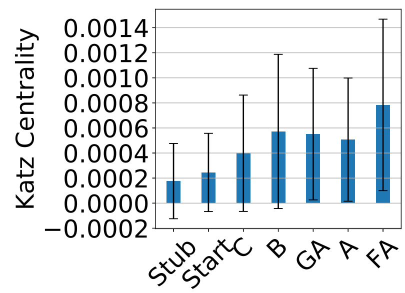

**(a)**



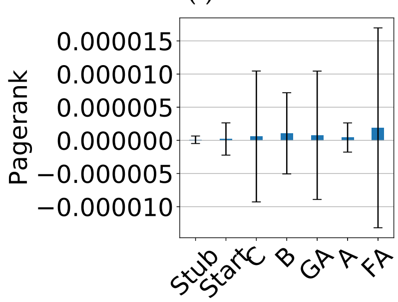

**(b)**



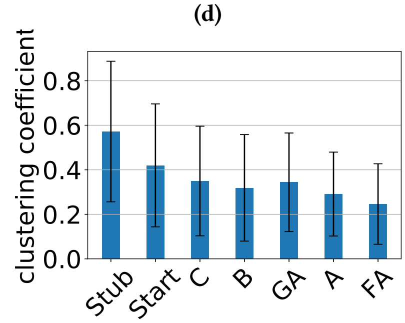

**(c)**



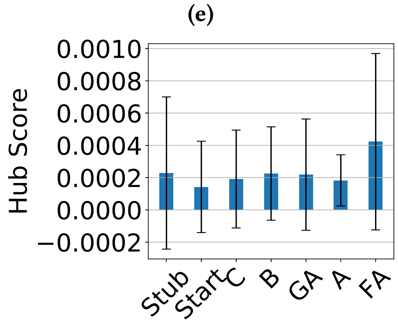

**(d)**



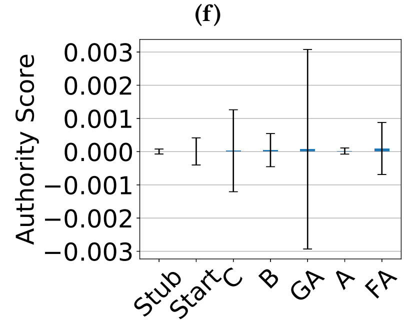

**(e)**



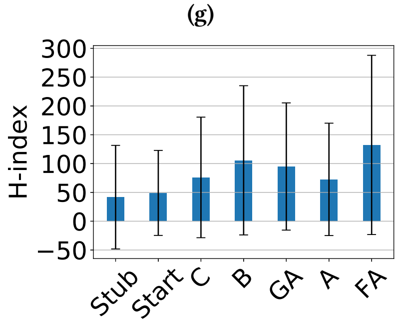

**(f)**



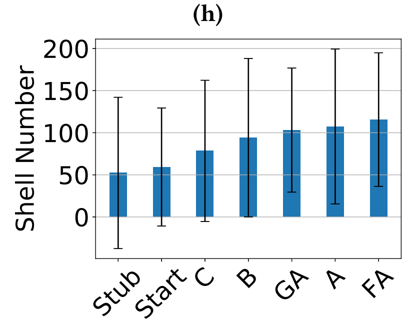

**(g)**




**(h)**




**(i)**




**(j)**

higher shells form the core of the network. There may be nodes in periphery that have a high degree; however, they are not densely connected to each other. Hence they are placed in lower shells. Therefore, the nodes with a high degree may not necessarily be a part of the core. Due to this reason, the core-periphery structure may provide additional information about the WikiGraph and its articles on top of the degree-based parameters.

We use the link structure of the entire Wikigraph for computing the node properties. For the quality grades estimation, we focus on the articles from seven quality grades,i.e., Start, stub, C, B, Good articles (GA), A and Featured Articles (FA), excluding the List and redirect articles due to their different article structure. This resulted in a total of 5, 722, 536 articles. We computed each of the above mentioned ten properties for these articles. The values of these properties with respect to the quality grades are shown in Figure 2.

It was observed that most of these properties varied with respect to the quality grade classes either in an increasing or decreasing fashion. The values for out-degree, Katz centrality, PageRank, hub score, and H-index generally varied in the increasing order while moving from stub to FA, while for clustering coefficient, they generally reduced. The reason is that due to a large number of neighbors of high-quality articles, the ratio of the number of links among these neighbors and the total possible links is likely to be small compared to the articles with fewer connections, thus resulting in a low clustering coefficient for high-quality articles. These variations indicate that the articles belonging to different quality grades typically exhibit a dissimilar network structure and thus substantiate the possibility of their usage for predicting the quality classes. It may be noted that we obtain a high standard deviation for the parameters due to a heavy long-tail distribution.

## 3.3 Building the predictive model

As the input data is highly imbalanced, with a smaller number of articles in higher-quality classes as compared to lower-quality classes such as start and stub (See Table 1), we pre-processed it to balance it. We used under-sampling using *NearMiss* [24] method to have the comparable number of articles in each class. We then built several classification models using methods such as SVM Classifier, Naive Bayes Classifier, Random Forest Classifier and KNN along with 10-fold cross-validation. We obtained the highest accuracy of 69.82% using the `Random Forest Classifier`. The values of the hyper-parameters were: `n_estimators` to be 180, `max_depth` to be 18 and `max_features` to be 1.0. Table 2 shows the confusion matrix of the prediction done using the Random Forest Classifier. It may be seen that even in the cases of misclassification, the misclassified classes are near to the true class in most cases. As an example, the misclassified cases of FA classes are predicted to be A, GA or B or C rather than the lower-quality classes. Table 3 shows the values of precision, recall and F1-score obtained for each class.

# 4 COMPARISON WITH EXISTING WORKS

We compare the efficiency of our model with the existing works in this direction. As our work provides prediction across the seven quality classes, we first compare with the works focusing on prediction across seven (or six as some existing works do not consider A-class) classes.

**Table 2: Confusion matrix obtained for Random Forest Model.**

| | FA | A | GA | B | C | Start | Stub |
|---|---|---|---|---|---|---|---|
| Actual FA | 525 | 51 | 74 | 7 | 2 | 0 | 0 |
| Actual A | 127 | 449 | 55 | 13 | 9 | 5 | 1 |
| Actual GA | 74 | 9 | 412 | 114 | 40 | 6 | 4 |
| Actual B | 7 | 2 | 75 | 429 | 112 | 27 | 7 |
| Actual C | 1 | 0 | 17 | 44 | 492 | 75 | 30 |
| Actual Start | 0 | 1 | 0 | 1 | 57 | 396 | 204 |
| Actual Stub | 0 | 0 | 0 | 1 | 22 | 118 | 518 |

One of the initial works in this direction used features including the number of headings, links, references, images, etc, and classified the articles into seven quality grades [28, 29] obtaining an accuracy of 58%. Another popular model is ORES [15] where the authors use features corresponding to the Wikipedia articles' content assessment[5]. They obtain an accuracy of 62.9%. Authors have also used neural networks on the document vectors of the articles [12] where they have obtained an accuracy of 55%. Recently, a work by Raman et al. [26] uses features obtained from the revision history networks of articles achieving 49.35%. Further, by merging the features thus obtained with the ORES features, they obtain an accuracy of 60.29%. In comparison with these existing works, our work focusing on the node-specific network properties obtains an accuracy of 69.82%. This shows the potential of the features obtained from the WikiGraph to estimate articles' quality. Moreover, by combining these features with non-network based features, we expect to achieve even higher accuracy, which is our study's future work.

**Table 3: Classification report for the Random Forest Model**

| Class | Precision | Recall | F1-score |
|---|---|---|---|
| FA | 0.71 | 0.79 | 0.75 |
| A | 0.87 | 0.68 | 0.76 |
| GA | 0.65 | 0.62 | 0.63 |
| B | 0.70 | 0.65 | 0.67 |
| C | 0.67 | 0.74 | 0.70 |
| Start | 0.63 | 0.60 | 0.61 |
| Stub | 0.67 | 0.78 | 0.72 |

Additionally, to compare our approach's efficiency with those performing binary classification, i.e., across high and low-quality articles, we also use our model to provide similar classifications and compare with the existing approaches as reported in Table 4. The articles have been divided into low and high-quality in different ways by the past works. For instance, some works compare FA articles with Start articles only [12, 32], others keep FA, GA a s high-quality articles and Stub to C as low-quality ones [12, 20], while a few others keep FA, GA as high-quality and Start to C as low quality, while keeping the stub articles out considering them

[5]https://en.wikipedia.org/wiki/Wikipedia:Content_assessment

**Table 4: Comparison with prior works on binary classes**

| Classifier | FA vs Start | (FA & GA) vs (C-Stub) | (FA& GA) vs (C-Start) |
|---|---|---|---|
| Lex at al. [20] | - | 84% | - |
| Wu et al. [31] | - | - | 66% |
| Xu et al. [32] | 84% | - | - |
| Dang et al. [12] | 99% | 86% | 90% |
| **Our approach** | **96.95%** | **98.76%** | **97.55%** |

basic enough for quality computation [12, 31]. As the Table 4 shows, the accuracy obtained by our method is higher even in the above cases of binary classification, further substantiating its efficacy.

## 5 CONCLUSION

The automated quality assessment of Wikipedia articles has been becoming a highly-active field of research. This study proposed a simple yet novel method for automatic quality grading of the articles based on the values derived from the network properties that any article holds as per its position in Wikigraph. Our model performs better than the existing models that use features based on either articles' content or revisions. Our study demonstrates that the structure obtained as per an article's placement in the WikiGraph network and how it connects with its neighbors is highly associated with its quality.

A method based only on network features is capable of performing comparably to content-based strategies due to the meaningful link structure of Wikipedia articles. The presence of a connection from an article $A$ to another article $B$ indicates the inclusion of details of $B$ into $A$. Further, many such connections certify the completeness or quality of article $A$, thus enabling the underlying link structure in revealing useful details about the articles' quality in general.

The automated identification of quality grades of Wikipedia articles may enable the editors to focus on content-editing and help the administrators easily find the articles that require attention. The future work of this study includes the usage of network properties of Wikipedia articles along with non-network properties, thus building a hybrid model to improve the accuracy of prediction further.

## ACKNOWLEDGMENTS

This work was supported in part by WOS-A, Department of Science and Technology, India [SR/WOS-A/ET-1058/2014] and CSRI, Department of Science and Technology, India [SR/CSRI/344/2016].

## REFERENCES


[1] Samy A Azer. 2015. Is Wikipedia a reliable learning resource for medical students? Evaluating respiratory topics. *Advances in physiology education* 39, 1 (2015), 5–14.

[2] Samy A Azer, Nourah M AlSwaidan, Lama A Alshwairikh, and Jumana M Al-Shammari. 2015. Accuracy and readability of cardiovascular entries on Wikipedia: are they reliable learning resources for medical students? *BMJ open* 5, 10 (2015).

[3] Grace Gimon Betancourt, Armando Segnine, Carlos Trabuco, Amira Rezgui, and Nicolas Jullien. 2016. Mining team characteristics to predict Wikipedia article quality. In *Proceedings of the 12th International Symposium on Open Collaboration*. 1–9.

[4] Joshua E Blumenstock. 2008. Size matters: word count as a measure of quality on wikipedia. In *Proceedings of the 17th international conference on World Wide Web*. ACM, 1095–1096.

[5] Ulrik Brandes, Patrick Kenis, Jürgen Lerner, and Denise van Raaij. 2009. Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th international conference on World wide web*. ACM, 731–740.

[6] Tibor Braun, Wolfgang Glänzel, and András Schubert. 2006. A Hirsch-type index for journals. *Scientometrics* 69, 1 (2006), 169–173.

[7] Luciana S Buriol, Carlos Castillo, Debora Donato, Stefano Leonardi, and Stefano Millozzi. 2006. Temporal analysis of the wikigraph. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE Computer Society, 45–51.

[8] Shai Carmi, Shlomo Havlin, Scott Kirkpatrick, Yuval Shavitt, and Eran Shir. 2007. A model of Internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences* 104, 27 (2007), 11150–11154.

[9] Anamika Chhabra and SRS Iyengar. 2020. Who Writes Wikipedia? An Investigation from the Perspective of Ortega and Newton Hypotheses. In *Proceedings of the 16th International Symposium on Open Collaboration*. 1–11.

[10] Anamika Chhabra and SR Sudarshan Iyengar. 2018. Characterizing the Triggering Phenomenon in Wikipedia. In *Proceedings of the 14th International Symposium on Open Collaboration*. ACM, 11.

[11] Anamika Chhabra, Rishemjit Kaur, and SRS Iyengar. 2020. Dynamics of Edit War Sequences in Wikipedia. In *Proceedings of the 16th International Symposium on Open Collaboration*. 1–10.

[12] Quang Vinh Dang and Claudia-Lavinia Ignat. 2016. Quality assessment of wikipedia articles without feature engineering. In *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*. 27–30.

[13] Raman Nuran et al. 2020. Classifying Wikipedia Article Quality With Revision History Networks. *Proceedings of Opensym* (2020).

[14] Linton C Freeman. 1977. A set of measures of centrality based on betweenness. *Sociometry* 1 (1977), 35–41.

[15] Aaron Halfaker and Dario Taraborelli. 2015. Artificial intelligence service 'ORES' gives Wikipedians X-ray specs to see through bad edits. *Wikimedia Blog* (2015).

[16] Meiqun Hu, Ee-Peng Lim, Aixin Sun, Hady Wirawan Lauw, and Ba-Quy Vuong. 2007. Measuring article quality in wikipedia: models and evaluation. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, 243–252.

[17] Leo Katz. 1953. A new status index derived from sociometric analysis. *Psychometrika* 18, 1 (1953), 39–43.

[18] Aniket Kittur and Robert E Kraut. 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 37–46.

[19] Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* 46, 5 (1999), 604–632.

[20] Elisabeth Lex, Michael Voelske, Marcelo Errecalde, Edgardo Ferretti, Leticia Cagnina, Christopher Horn, Benno Stein, and Michael Granitzer. 2012. Measuring the quality of web content using factual information. In *Proceedings of the 2nd joint WICOW/AIRWeb workshop on web quality*. 7–10.

[21] Xinyi Li, Jintao Tang, Ting Wang, Zhunchen Luo, and Maarten De Rijke. 2015. Automatically assessing Wikipedia article quality by exploiting article–editor networks. In *European Conference on Information Retrieval*. Springer, 574–580.

[22] Andrew Lih. 2004. Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. *Nature* 3, 1 (2004), 1–31.

[23] Jun Liu and Sudha Ram. 2018. Using big data and network analysis to understand Wikipedia article quality. *Data & Knowledge Engineering* 115 (2018), 80–93.

[24] Inderjeet Mani and I Zhang. 2003. kNN approach to unbalanced data distributions: a case study involving information extraction. In *Proceedings of workshop on learning from imbalanced datasets*, Vol. 126. ICML United States.

[25] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.

[26] Narun Raman, Nathaniel Sauerberg, Jonah Fisher, and Sneha Narayan. 2020. Classifying Wikipedia Article Quality With Revision History Networks. In *Proceedings of the 16th International Symposium on Open Collaboration*. 1–7.

[27] Besiki Stvilia, Michael B Twidale, Linda C Smith, and Les Gasser. 2005. Assessing Information Quality of a Community-Based Encyclopedia.. In *IQ*.

[28] Morten Warncke-Wang, Vladislav R Ayukaev, Brent Hecht, and Loren G Terveen. 2015. The success and failure of quality improvement projects in peer production communities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 743–756.

[29] Morten Warncke-Wang, Dan Cosley, and John Riedl. 2013. Tell me more: an actionable quality model for Wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration*. 1–10.

[30] Thomas Wöhner and Ralf Peters. 2009. Assessing the quality of Wikipedia articles with lifecycle based metrics. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*. ACM, 16.

[31] Guangyu Wu, Martin Harrigan, and Pádraig Cunningham. 2012. Classifying Wikipedia articles using network motif counts and ratios. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*. 1–10.

[32] Yanxiang Xu and Tiejian Luo. 2011. Measuring article quality in Wikipedia: Lexical clue model. In *2011 3rd Symposium on Web Society*. IEEE, 141–146.

[33] Haifeng Zhang, Yuqin Ren, and Robert E Kraut. 2020. Mining and Predicting Temporal Patterns in the Quality Evolution of Wikipedia Articles.. In *HICSS*. 1–10.

[34] Vinko Zlatić, Miran Božičević, Hrvoje Štefančić, and Mladen Domazet. 2006. Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E* 74, 1 (2006), 016115.